DOCUMENT RESUME

ED 124 243 JC 760 324

AUTHOR Nelson, Dennis E.
TITLE Formulating Instrumentation for Student Assessment of Systematic Instruction.
INSTITUTION College of the Mainland, Texas City, Tex.
PUB DATE 19 Mar 76
NOTE 34p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Community Colleges; Course Content; Evaluation Criteria; *Evaluation Methods; *Junior Colleges; *Literature Reviews; Questionnaires; *Rating Scales; State of the Art Reviews; *Student Evaluation of Teacher Performance
IDENTIFIERS College of the Mainland; SASI; *Student Assessment of Systematic Instruction

ABSTRACT

A review of the literature indicates the recent popularity of student evaluation of faculty in (1) providing feedback to faculty so that they can modify or improve their instruction; (2) acquiring information to be used in evaluating faculty, and in making tenure and promotion decisions; (3) acquiring information to be disseminated to students so that they can make course and curriculum choices. In spite of the high number of evaluation instruments produced in recent years, few have specifically been geared to the theoretical model of instruction being applied in the course to be evaluated. In order to alleviate this problem, a new student evaluation instrument--The Student Assessment of Systematic Instruction (SASI)is being developed at College of the Mainland. The proposed instrument divides the evaluation into six categories: the organization and structure of learning, the quality of learning materials, students' perception of the value of course content, students' perception of their own personal growth, quality of teaching, and quality of student-teacher interaction. Students are asked to rate the section in which they are enrolled in relation to other sections in the same course, department, division, and college; and to explain what reasons they have for their rating. Results will be reported by computer, and a sample computer printout is appended. (Author/NHM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED124243

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FORMULATING INSTRUMENTATION

FOR

STUDENT ASSESSMENT OF SYSTEMATIC INSTRUCTION

Dennis E. Nelson, Ph.D.
Resident Consultant in Educational Evaluation

College of the Mainland
Texas City, Texas

March 19, 1976

JC 760 324

FORMULATING INSTRUMENTATION FOR STUDENT ASSESSMENT OF SYSTEMATIC INSTRUCTION

HISTORICAL USES FOR STUDENT EVALUATION

The notion that it might in some way be useful, or at least interesting, to seek out information from students concerning what they think about the current instructional experiences in which they are involved is far from new. Centuries ago such data was not only sought out, but was in some instances used for purposes which involved the levying of fines upon faculty members (Rashdall, 1936). At American universities, while the results were not as severe, student ratings of teachers were being collected in excess of 50 years ago (Centra, 1972). Rather extensive histories of the student evaluation movement have been compiled elsewhere (Werdell, 1967, Ebel, 1970). The purposes for gathering student evaluations have ranged from pure intellectual curiousity to the aforementioned seeking of evidence to be used in meting out subsequent punishment to dilatory instructors. For the most part however, categories of purposes in accumulating student ratings of instruction can be reduced to three. They involve 1) provision of feedback to faculty with a view toward modification and improvement of instruction, 2) acquistion of information to be used to evaluate faculty in some manner and to make decisions with regard to tenure, promotion and similar personnel concerns, and 3) dissimination to students of data thought to be helpful as they make choices viz a viz various course offerings and curriculum options (McKeachie, et al., 1975).

It has been pointed out elsewhere (Ronan, 1973) that students are certainly in a position to observe at close range a large quantity of teaching and are in a good position to provide some reactions to what they experience. This certainly accounts, at least partially, for the wide spread investigation of student ratings. A few hours spent in a perusal of publication abstracts from the past fifteen years or so will reveal to the reviewer the upsurge in interest with regard to the topic. Though there is evidence to support the idea of at least a temporary decline in the use of student ratings to evaluate faculty (Gustad, 1961, 1967) the overall trend seems to foreshadow a trend of increased use of student ratings for all three purpose categories. The growing scarcity of financial resources for education, the concurrent pressure for accountability and similar phenomena are no doubt contributing to this increase. Strong advocacy for the gathering and use of student ratings has also emerged from the students themselves (Werdell, 1967), and even from faculty (AAUP, 1974).

Though often greeted with initial hostility by many faculty members, especially when the formulation strategy of the device did not include their input or where the purposes to which the data will be put are unclear, or are viewed as unacceptable, student ratings will most likely be a part of life in higher education for some considerable time to come. In the majority of instances, faculty response probably will follow to a large degree, the pattern of initial hostility, gradual acceptance, and eventual support found by Slobin and Nichols (1969). Costin, Greenough, and Menges (1971) provide a concise summary of the arguments set forth by both the supporters and opponents of student ratings, particularly with respect to their proposed role in faculty related decisions by the administration.

PROPOSED APPROACHES

The approaches used to collect student reactions to instructional experiences, have taken widely divergent paths and the important relationship between the purpose to which the data has to be put, and the approach used in developing the instrument has not always been clearly thought out. Most instruments have been the result of student groups, individual faculty members or departmental committees. On far fewer occasions have the devices been the result of concerted work by specialists in educational measurement. Overwhelmingly they have been predominately teacher-oriented and have tactically assumed a very narrow view of the instructional process. It is reported that some of the earliest American studies of student ratings, done in the 1920's by Edwin Guthrie, were accomplished by asking students in a straight forward manner to describe what they considered to be the salient qualities of good teaching. Content analysis was then employed to reduce this raw data to a format useable for presentation to subsequent groups of students who indicated which characteristics were applicable to the best teacher with whom each had studied (French-Lazovik, 1974). Eventually, a scale was developed from this early work which proved to be the central technique of measurement used in Guthrie's later research on student ratings (Guthrie, 1954). A variant to this strategy utilizing critical incidents has also been attempted (Owen, 1967). While this general strategy has been extremely influential as judged by its frequent use in other investigations (Smith, 1944, Crawford and Bradshaw, 1968), it has received criticism on the grounds that it provides insufficient guarantees that all relevant variables have been included and that it fails to indicate the relative importance of each characteristic perceived as being involved in good teaching (French-Lazovik, 1974).

The notion of relying on student self-reflections in order to discover what they supposedly value in teaching was avoided by Coffman (1954) when he simply factor analyzed a student rating device in use with an aim toward revealing the central factors which had resulted in the student ratings obtained. While such an approach most certainly spoke to the inadequacies ascribed to the Guthrie strategy, the particular device analyzed contained relatively few items and was rather restricted in the scope of the instructionally related topics covered. Additionally, French-Lazovik (1974) suggests that the application of a prediction model would have proven much more useful since it would have yielded information about how well students' estimates of teaching effectiveness could in fact be predicted as well as revealing the best predictors and their relative impact.

A number of investigators have attempted through factor analytic techniques to survey a wide range of instruments and items in order to retrieve the essence of student ratings of instruction. In one of the finest such efforts McKeachie et al. (1964) reviewed a large number of devices which had been used to collect student ratings of courses and instructors and through factor analysis attempted to find factors that would be consistent across time as well as across different students and different teachers. Kaiser's method of studying factor similarities yielded six factors meeting such criteria. Skill, overload (relating to student work), structure, feedback, group interaction, and student teacher rapport were the labels assigned to the resultant constructs. Similarly Hildebrand, Wilson, and Dienst (1971) used a large item pool and through discriminant analysis sought a few factors which would prove to be the central contributors to the wide divergence of

student ratings. However methodological problems plague such approaches to the problem, whether or not a prediction model is used as the focus of the study. For example, large numbers of variables are involved, and there is substantial disagreement among knowledgeable professionals not only as to the ideal size and composition of the item pools to be examined but with regard to the most suitable techniques for subsequent data analysis as well.

Having examined some of the methodological "pluses" and "minuses" of each investigation, and subsequently searching for commonalities among a number of factor analytic studies which attempt to account for extent variance in student ratings, French-Lazovik (1974) concludes that students seem to rate highly instructors who tend to be clear in setting forth what is to be learned, who in some unspecified manner bring to life or broaden student interests, and who bring about increased intellectual activity and higher general motivation in students. It would seem that such results could be achieved by differing instructional approaches, and by faculty members who differ in style or personality variables.

Silberman and Allender (1974) assert that a more helpful course of action at least from the standpoint of providing feedback to instructors would be to use a semi-projective technique. Their device consists of a written response from students of some fifteen minutes duration based on the hypothetical context of telling a prospective student about the course. No constraints or guidelines are given to influence the direction or content areas of the student responses and content analysis procedures are used to quantify the results which are divided into scores for "evaluative tone" and "input".

This latter score relates to a personal growth dimension which has so often been seen as being a very desirable but elusive variable on which to gain solid evidence. The desire of institutions to show some positive change in student personal growth as a result of organized educational experiences appears in a device developed by Hartley and Hogan (1972) in which this variable is the concern of half the items. There would seem to be no appreciable opposition to gaining such evidence from student ratings if it in fact exists, and this instrument is notable for such an attempt.

Yet another organizational pattern is exemplified by a device designed to fit a wide variety of course offerings and organized on what would seem to be a "logical elements" approach. The FACE (Sorkloff and Doubler, 1971) contains items aimed at tapping six areas of course related stimuli: The Professor; Text and Readings; Examinations; Papers and Reports; Recitation; and Laboratory. Not only can the instructor select areas of the instrument which are most applicable to a particular course, but the FACE is noteworthy in its attempt to move away from an almost exclusive emphasis on instructor characteristics and behavior found in the vast majority of devices used to solicit student ratings.

Concomitant with the increasing interest in and use of student ratings has been an increased interest on the part of commercial publishers in the development and distribution of instruments designed to be both usable and useful in a wide variety of institutions and instructional contexts. Perhaps among the best known and most widely used is the Student Instructional Report (SIR) (Centra, 1972). Shaped partially by facutly review of proposed items

as well as upon instructional "factors" found in previous research, this popular instrument seeks student responses along a four point scale ranging from "Strongly Agree" to "Strongly Disagree" with an additional category of "Not Applicable or Don't Know". A few items are formatted on a slightly different scale. Content areas include Course Organization and Content, Instructor-Student Relations, Communications, Assignments and Evaluation, Lecture, Discussion-Seminars, Laboratories, and Student Involvement, in addition to respondent background data. A comments section is also included which encourages open-ended written responses on any of the above topics, and an opportunity is also provided for student critique of the device itself. Advantages of this instrument are its wide applicability and its focus on a clear purpose: feedback to faculty for course improvement. Yet there are inherent limitations in a device with "something for everyone" and it is likewise not uncommon for considerations related to projected commercial success and wide market applicability to overshadow such concerns as theoretical clarity and the uniqueness of particular contexts of use. In addition, it is not clear whether the use of faculty to screen items and remove those of which they did not approve, contributed to strength or weakness in the device.

UNDERLYING CONCERNS OF EVALUATION

Regardless of advantages and disadvantages connected with the particular approach and format of data collection efforts or the philosophical viewpoint represented in their mode of development and construction, there are underlying concerns in connection with virtually every student evaluation that remain the same.

Among these are the stability or reliability of ratings and their validity. Correlations between ratings made at intervals ranging from two weeks (Lovell and Haner, 1955) and half a semester (Costin 1968) to a year (Guthrie, 1954) have consistently produced correlations of .70 and higher. Several different instruments have also been found to have considerable internal consistency (Spencer and Aleamoni, 1970; Harvey and Barker, 1970; Guthrie, 1954; Maslow and Zimmerman, 1956). Regardless of the particular method utilized to calculate reliability, there is considerable cause for stating that students can rate instruction with substantial consistency. The issue of validity, however, raises a number of difficulties with regard to what evidence is accepted as being sufficient. McKeachie (1969) has observed that faculty members could judge the validity of student ratings by matching them against the goals the instructor held for the course. In what has been termed a "student as consumer" orientation, evidence exists that students tend to give higher instructor ratings in relation to the extent that they perceive they have met course objectives (Lathrop, 1968). This way of viewing validity would also apply to further course enrollment by students who have given high ratings. McKeachie and Solomon (1958) report some evidence of such a relationship though the evidence is not always consistent (Tobias and Hanlon, 1974).

An alternative source of evidence regarding the validity of student ratings is found in the ratings given instructors or particular courses by supervisory personnel or other colleagues. Though significant correlations between student and supervisory ratings are not always found (Webb and Nolan, 1955), most positive results have been sufficiently low in magnitude to suggest

that student ratings are substantially different in nature and have their own contribution to make in attempts to evaluate instructional experiences (Costin, Greenough, Menges, 1971). It might be argued of course that the variable of experience and training is the variable to be examined rather than supervisory or colleague ratings. Findings from such studies, however, range from negative correlations (Rayder, 1968) to no significant relationship (Heilman and Armentrout, 1939), to a positive correlation between experience and ratings (Walker, 1969) and between academic rank and ratings (Gage, 1961). The findings of these and other similar studies can however be understood more clearly as an analysis is made of the particular definitions of "experience" and "training" which are used. Substantial positive relationships can be shown between specific aspects of student ratings and precisely defined measures of training or experience. Costin (1968), for example, found that student ratings of graduate assistants on subtopics as "feedback" and "group interaction" were clearly related to whether or not the assistants had completed a course emphasizing instructional techniques applied to the subject area in which they were teaching.

Perhaps the most intensely examined source of potential validity for student ratings has been the realm of performance. The notion is that if student ratings of the instructional experience are quite positive then it is reasonable to expect that a greater quality or quantity of learning, or a better academic product will result. The active debate over some studies of this nature which have been undertaken is no doubt related in part to the technical difficulties which accompany the execution of such studies (Cronbach and Furby, 1970). For example, (Rodin and Rodin, 1972) reported a negative

relationship between achievement and student ratings but the study has been widely criticized. Methodological flaws may well account for the nature of the findings (Rodin, Frey, and Gessner, 1975). As Eagle (1975) points out, not only have subsequent studies failed to replicate the findings, but investigators such as Bryson (1974), Frey (1974), and Gessner (1973) have all found achievement-rating correlations to be positive in nature. It should be clearly noted, however, that finding a link between performance or achievement and student ratings is quite different from finding such a link between ratings and the biasing variable of grades. Relationships of the latter sort, have been weak in magnitude (Costin, Greenough, and Menges, 1971). In a very well-executed attempt to determine if the positive relationships between achievement and ratings found in earlier studies (Frey, 1973; Gessner, 1973; Sullivan and Skanes, 1974) were reliable, Frey, Leonard, and Beatty (1975) produce evidence which indicates that stable links can be established between a number of aspects found in student ratings and subsequent measured achievement. This "directional thrust" as Scott (1975) phrases it, is the most significant facet of the study. Since one barrier to obtaining strong consistent validating evidence for student ratings has been the inadequacy of the instruments used to measure student learning, Scott urges further effort toward criterion measures which would allow a more complete examination of the relationship between achievement and ratings. One possibility suggested in this regard is the use of matrix item sampling (Sirotnik, 1974).

The central issue in the quest for the establishment of validity for student ratings is that highlighted by Crittendon and Norr (1975). Ratings must be distinquished from variables which act as "biasing factors". Class size, student sex, major, ethnic group, and a host of other influences must be found to be substantially independent of ratings. Useful references to a myriad of such investigations are readily available (Crittendon and Norr, 1975; Menges, 1972; Centra, 1972; Costin, Greenough, and Menges, 1971). When this massive amount of data is surveyed it is in some ways remarkable and as Eagle (1975) notes "encouraging" that the correlations are for the most part relatively low. These so-called "biasing factors" simply do not account for the majority of the variance found in student ratings. It would seem that information about student perceptions of instructional experiences is sufficiently independent from other sources of evaluative information to be one of a number of useful "windows" (Scott, 1975), through which to view instruction. Similarily it appears that the information obtained can be both reliable and valid, and that such data can have a useful impact on students, faculty, and administrators (Centra, 1973). Each instrument must, however, overcome the challenges of issues related to validity and reliability for no blanket validity exists (Eagle, 1975).

RATIONALE FOR ADDITIONAL INSTRUMENTATION

Given the abundance of a wide variety of student rating forms (Werdell, 1967; Ebel, 1970 provide a survey of many currently in use) it seems on the face of things to be an ill advised task to formulate yet another instrument. Perhaps the key to understanding the initiation of work on instrumentation related to student ratings at College of the Mainland lies in a suggestion by Halstead (1970) that rating devices should be geared to the specific theoretical model of instruction being applied in the courses to be evaluated.

This suggestion has, it appears, gone unnoticed. Current instruments seem to assume that instruction consists of a series of lectures which accompany limited discussion, and/or a laboratory experience of some form. Such devices are almost exclusively formulated on the basis of teacher behavior or inferred traits and reflect a highly instructor-dependent model of learning. Additionally, the possibility of providing computer generated results, including comparisons with other horizontal or vertical categories of instruction within or outside an institution, as well as indicating in a like manner possible sources of assistance for faculty revision of instructional experience, has scarcely been acknowledged. Due to these and similar reasons, and as the outgrowth of an institutional commitment to programmatic solicitation and utilization of student opinion, the present project was undertaken.

DEVELOPMENTAL GUIDELINES

In harmony with general systems thinking as applied to educational contexts, an examination of the current state of collecting and using student reactions to instruction was conducted. This effort yielded the knowledge that while there were a number of quasi-formal student reaction devices in use, they were of highly uneven quality, usually containing serious flaws from a measurement point of view and often, the kind of information that was obtained could as easily have been gained from a supplemental section added by the instructor to another device. There were however at least two exceptions to this general state of affairs and these two instruments will prove useful in latter stages of developing an institution wide program of assessing student perceptions of instruction. One of these devices, currently

being used with entering students suffering from severe deficits in terms of basic skills, might well become an alternate form of the instrument described herein. On a college wide level, the form most widely used underwent the scrutiny of both internal and external evaluators. The conclusions at which both individuals arrived were in essential agreement (Dobbins, 1975), and reflected the areas of inadequacy outlined above together with others of a more technical nature. To further shape the direction of the instrumentation effort, several administrative leaders within the institution who were directly concerned with instruction provided a number of constraints upon the device and a series of desirable specifications to be met. In summary the expressed desire was that the final product be

- based in and reflective of the institutional commitment to what would be termed systematic instruction. This model of instruction, adopted and fostered by the college, involves the use of concepts such as student preassessment, clear objectives across all categories and levels of the cognitive and affective domains, objectives tied closely to behavior, flexibility and ready availability of learning materials, continious progress and variability in instructional pace, and non-punitive grading directly related to course objectives

- applicable to all divisions within the college including academic, vocational-technical and non-credit offerings

- appropriate for administration in quasi-traditional courses and should provide some indication of the degree to which a course is moving toward the model of instruction encouraged by the college

- feasibly administered at any time during the duration of a course. (Many courses at the college involve continuous registration and course length varies widely as more dimensions if instructional flexibility are introduced

- administered within 30 minutes or less and involve a minimum amount of faculty time relative to administration

- scored by computer and returned to the faculty in time for results to influence the subsequent offering of a course
- amenable to faculty self interpretations with a minimum reliance and assistance from outside the institution
- flexible enough to allow for collection of additional information by individual faculty
- designed so as to provide comparisons between instructors, sections of given course, course groupings within a division (departments), and divisions within the college as well as comparisons between a particular course section and each of these instructional levels.
- capable of measuring student preceptions of personal growth and an emerging "independent learner" orientation which has been a high priority outcome sought by the college since its inception.

SELECTION OF FACTORS AND ITEMS FOR FIELD TEST

Provided with these product specifications and some notion of what has been done institutionally with respect to student perceptions of instruction, developmental work was begun within the resource constraints allowed in terms of time, personnel, and finances, Factors which were selected as the result of an examination of literature related to individualized instruction and other systematic approaches. These "factors" included 1) the organization and structure of the learning experience, 2) the learning materials that are utilized, 3) the perceived value of the course content, 4) the instructor or learning manager as a teacher, and 5) the quality of the interactions which take place between the students and the instructor or learning manager. Many of the items chosen to represent these factors as the initial version of the instrument were those found to have been highly associated with constraints

of a like nature in research with other student rating devices. Preliminary statistical studies show these items to be reasonably stable. Others of course, represent an attempt to find additional items which will serve to insure a statistically acceptable minimum number of items for each factor.

A number of psychological advantages to the individual have been claimed to accrue to students exposed to flexible systematic approaches to instruction (Young and Van Mondfrans, 1972). Due to such claims and an institutional stress upon promoting student growth through instruction, a number of experimental items were included which relate to a hypothesized factor of "perceived personal growth." This factor is at present a hypothetical notion and evidence may or may not be found to support its presence. It may in fact be related to the dimension of student involvement or personal needs as found in some previous studies or highly similar to a scale of internal-external locus of control. Additionally, two items which solicit information of importance to a wide variety of personnel at the institution are present. These relate to a perceived incidence of cheating and a student analysis of the evaluation activity itself. The latter is considered of high priority in College of the Mainland documents relating to systematic course development procedures to be used at the institution.

The primitive edition of the instrument now being field tested for subsequent editing and revision also contains an abnormally large number of items related to student background characteristics. While some of this information will yield helpful additions to the currently held profile of the college student body, especially with regard to those enrolled in non-credit offerings, most

of these will be useful in confirming the relationship or lack thereof between ratings and numerous biasing variables and will be made use of in subsequent validity studies connected with the instruments refinement. Since its avowed purpose is to tap student perception of where a particular course stands as currently constituted with respect to certain dimensions thought to be an integral part of the model of instruction adopted by the college it has been christened, Student Assessment of Systematic Instruction (SASI).

SASI: CURRENT CONTENT AND FORMAT

The specific items grouped according to the factors which constitute the field test version of SASI are listed below. It should be noted that as they appear on the student rating form itself, approximately half the items associated with each factor are reversed in direction to avoid the effects of response set or inattention and all factors are systematically rotated with no two consecutive items representing the same factor.

I. ORGANIZATION & STRUCTURE OF LEARNING

1. What I am to learn is made clear.
2. Assignments are clear and definite.
3. I understand how I am being graded.
4. I am graded on the basis of what I am expected to learn.
5. Classes meet and are dismissed on time.
6. The amount of work assigned is about right.
7. The time allowed to cover the material is about right.
8. I received a printed course document which is useful.
9. The course is well organized.
10. The course is following the printed course document.

II. LEARNING MATERIALS

1. Printed materials are readable.
2. The textbook is satisfactory.
3. I like the materials used.
4. The audio-visual materials for this course are helpful.
5. Students have a choice of learning materials.
6. Necessary materials are usually available.
7. The learning materials are helpful.
8. Materials are understandable.
9. The library is helpful to me in this course.
10. Materials seem to be presented on my level.

III. PERCEIVED VALUE OF CONTENT

1. What I am learning is worthwhile.
2. Important topics are stressed.
3. I have become interested in further study of the subject.
4. I can use what I am learning.
5. I would rate this course higher than others I have taken here.
6. Taking this course has been a good experience.
7. I attend this class regularly.
8. I would rate this course higher than most I have taken.
9. Most people would benefit from this course.

IV. STUDENT PERCEPTION OF PERSONAL GROWTH

1. This course is changing my view of myself.
2. I am developing a better understanding of other people.
3. This course is helping me rely more on myself.
4. I feel closer to other people as a result of this course.
5. I now know more about myself.
6. I will now probably study more on my own.
7. This course is changing how I think.
8. It is easier now to see other points of view.
9. I am learning that I can change the way things are.

V. CLASS LEADER OR MANAGER AS A TEACHER

1. Overall, the instructor is a good teacher.
2. The instructor seems to be interested in the subject.
3. The instructor speaks clearly.
4. The instructor answers questions.
5. The instructor knows a great deal about the subject.
6. The instructor is sure and confident.
7. The instructor is usually free from annoying mannerisms.
8. The instructor keeps my interest.
9. I would like to take another course from this instructor.
10. The instructor's answers to questions are accurate.
11. The instructor can handle difficult situations.

VI. QUALITY OF STUDENT INTERACTION WITH LEARNING MANAGER OR TEACHER

1. The instructor keeps regular office hours.
2. I can get help from the instructor outside of class.
3. The instructor shows respect toward students.
4. The instructor takes an interest in students as persons.
5. Students can express their opinion.
6. Differences of opinion are allowed.
7. Grading is impartial.
8. There is quite a bit of student participation in the class.
9. The instructor seems to like being with the students.
10. No one in this course gets special treatment.

VII. INDEPENDENT ITEMS

1. There is no cheating.
2. Having students evaluate courses is worthwhile.
3. Supplementary questions prepared by instructors.

In terms of individual data, the student is asked to indicate:

1. Type of program in which he is enrolled.
2. Current enrollment pattern with respect to credit and non-credit courses.
3. Number of semesters of prior attendance at the institution.
4. Extent and type of current employment.
5. Characteristic level of perceived academic performance.
6. Ethnic group.
7. The number of pre-school children at home.
8. Birthdate.
9. Sex.
10. Previous number of times this semester the SASI has been filled out.
11. Language commonly spoken at home.

WITH RESPECT TO EACH ITEM, THE STUDENT IS ASKED TO READ EACH STATEMENT CAREFULLY AND DARKEN THE SPACE ON THE ANSWER SHEET THAT BEST DESCRIBES HIS OR HER OPINION ACCORDING TO THE FOLLOWING SCALE:

(1) Strongly Agree. You strongly agree with the statement as it applies to this course.
(2) Agree. You agree more than you disagree with the statement as it applies to this course.
(3) Disagree. You disagree more than you agree with the statement as it applies to this course.
(4) Strongly Disagree. You strongly disagree with the statement as it applies to this course.
(5) Not Applicable or Don't Know. The statement does not apply to this course, or you simply are not able to give a knowledgeable response.

STEPS OF DATA ANALYSIS

Two levels of analysis are performed for each of the item groupings included in the instrument. The first level of analysis is designed to answer the question "With respect to each factor measured, how does this section as a whole compare in relation to other sections of the same course, the same department, the same division, and in the college?" The answer to the question is given through use of one of the following terms: high, above average, average, below average, or low. The computation procedure utilized to execute the first level of analysis can be reduced to the following steps:

1. Determine the common variance of the items in an item group.

2. Establish weighting for each of the items in the item group based on the degree of common variance of the item with the other items (some items are better measures and therefore will carry higher weight).

3. Compute the group statistics for the item group -- mean (or average), standard deviation, standard error of the mean -- computations performed by class, course, department, division, and college.

4. Since the arithmetic mean (or average) is subject to errors of measurement, calculate the confidence interval about the mean such that we can state with 95% confidence that the "true" mean lies within the interval.

5. Determine the item group rating according to the following rules:

   a. Having established the confidence interval for a given class, rate the class below average compared to college, the course, department, division, if the mean of the designated grouping falls within the confidence interval of the class; and rate the class above average if the mean of the designated grouping falls below the lower limit of the class confidence interval.

b. If the class is rated above average and the class mean is 10* or more points above the mean of the designated group, revise the rating of the class to "high;" or if the class is rated below average and the class mean is 10* or more points below the mean of the designated group, revise the rating given the class to "low."

*Note: All scores are converted to standard scores with a mean of 50 and a standard deviation of 10 for statistical reasons. Hence, 10 points is 1 standard deviation unit.

The second level of analysis performed for each of the item groupings is designed to answer the question -- "how may the rating obtained in the previous analysis be explained?" Our analytical methods can give only a partical answer, or perhaps only some clues. The answer(s) are framed in terms of item ratings given in the factor by factor report for each course-section.

The items which form the basis of the group score and rating are themselves ranked high to low based on the mean student ratings received by each. The weighting of each item is displayed along with the arithmetic average and standard deviation of the ratings received. Items receiving high mean ratings by students and those possessing high weightings with regard to a particular factor are most influential in producing high group scores on that factor. Items having low mean ratings, yet possessing high factor weights, contribute to low group scores. Lower weighted items, irrespective of mean ratings, contribute less to the group score.

REPORTING OF RESULTS

A report in the form of a multi-page computer printout is subsequently prepared for each faculty member which provides three basic kinds of information. First, student perceptions relative to each factor in every course-

section for which the instructor is responsible are summarized and reported. Second, a number of relative comparisons are made between student perceptions in a particular class with perceptions of other section of the course, the division and the institution. Third, an item by item breakdown is given with respect to student responses on every item associated with each factor. Those items which do not contribute significantly to any of the six distinct factors currently being measured are grouped together at the end of the report under the heading of "Independent Items" with only item information being reported since no known factor is involved.

Appendix A contains a page from a computer report typical of that received by each faculty member. Note that in the top left portion of the page many of the symbols used in the report are explained. To the right of that explanation is found a table which summarizes the number of respondents to the instrument in terms of classes, courses, students, and the like.

The remainder of the printout is presented on a course section by course section basis with the results (by factor) for each course section being reported in the following order: (1) Structure and Organization of Learning, (2) Learning Materials, (3) Value of Course Content, (4) Perceived Personal Growth, (5) Instructor or Learning Manager as a Teacher, and (6) Quality of Student-Instructor Interaction.

With regard to each factor, results are presented in an identical sequence. On the left, the class, course, department, division, and college mean ratings for the factor in question are listed along with numeric limits within which we can be confident (at a 95% level) that the "true" mean rating lies. This concept of confidence limits, though perhaps initially difficult

to grasp, is an important one. All measurements contain some error due to imperfect instrumentation, response variables, and a plethora of other influences. Even repeated measurements with a refined instrument under highly controlled conditions would not eliminate all sources of error. Any scores or measurements we present then only approximate a true and accurate measurement. Therefore, an attempt has been made to report a range within which we can be confident that the "true" score is located. Additionally, the standard deviation, which is an indication of how dispersed or spread out the ratings are, is provided. In this example, note that the ratings of the class tended to be more spread out or vary to a greater extent than did student perceptions of the same factor generally throughout the division.

To the right of this data, a graphic display of comparative arithmetic means and confidence limits is given with these same relationships summarized verbally on the right hand side of the page. One can easily see, for example, that the mean class perception on the factor summarized in this portion of the report is, with respect to the mean perceptions in all sections of the same course below average but is above average when compared with perceptions of the same factor throughout the department.

Detailed data with respect to each item associated with the particular factor being reported, follows. At the extreme left of this display the "Rate" column is indicative of whether the average student rating of the item is significantly above, near, or below average. The items average rating on the four point response scale being utilized follows, along with a measure of how dispersed or spread out the ratings were on this item (SD).

Subsequent columns in this display indicate (1) the relative importance or impact of the item in the factors composition, (2) the percentage of respondents checking the "not applicable/don't know" option, (3) a classification of this percentage as being significantly above, near, or below average, and (4) the number of the item as it appeared in the instrument.

Items are listed in an order reflecting their positive contribution to the factor rating. Examining item weightings and printed order visually thus provides a simple and rapid method for discerning potential reasons for obtained factor ratings. The same data might be utilized as initial guideposts in determining likely strategies for increasing student ratings on a particular factor. Note that items which were originally phrased in negative terms have, in the reported results, been reversed and are stated in the positive.

Item information includes an analysis of the degree of occurrence of missing responses and responses to the "not applicable" response alternative. Overall response rates are computed by item for the college as a whole and comparison is then made to the item response rates for each class. Differences are indicated when the number of missing and "not applicable" responses differs significantly in the class.

Additional information concerning the number of occurrences of "no response" and "not applicable" responses is provided with the item display which contains the mean rating and standard deviation. Items will be excluded from reports for a particular class analysis when the number of responses indicating "not applicable/don't know" exceeds 50% of the total number of students in the class who fill out the instrument.

CURRENT FIELD TEST

The instrument as herein described was administered to a non-representative sample of the student body during December, 1975. Instructors administered the form on a voluntary basis but all instructors were encouraged to utilized it with at least one of their classes. It is estimated that between ten and fifteen percent of the fall student enrollement responded at least once to the device but at time of administration owing to the degree on instructional flexibility already extant in the normal operations of the college, many courses had already concluded, and a number of additional students had already completed courses that were, for other students, still operational. While instructors were encouraged to supplement the device with items covering concerns unique to a particular class or course, most gained such data by using whatever devise they had employed previously. Therefore a number of students responded to more than one device in evaluating a particular course. Semi-structured or open ended item additions were also encouraged. Anecdotal reports indicated an average response time of approximately 25 minutes for initial responders with a response time approaching 15 minutes for some students who completed the form in a fourth or fifth class. Owing to a number of variables adversely effecting the fall administration, including sample composition, the unfortunate timing of materials distribution including a delay in placing a user's manual (Nelson and Dobbins, 1975) in the hands of faculty, a decision was made to process information from a far more extensive administration of the SASI in March of 1976, prior to revising and refining the student form and data collection procedures.

FUTURE DIRECTIONS

The time and effort as well as the other resources needed to develop and validate a most measurement instrument is substantially indeed. College of the Mainland is committed to using information from students as one of several inputs through which to evaluate and improve instructional experiences. During the current academic year a beginning has been made. As one result of the student assessments collected this year, each item will be scrutinized for possible retention in its present or an altered form or replaced. In a similar vein, the existing factors included in the field-test version of the device will be examined for possible changes in number and structure. A number of approaches will be investigated in a quest for validity.

With a somewhat more extended time frame in mind, the possibility of using computer reports to provide more detailed diagnosis of the aspects of instruction needing modification, and as a referral mechanism for individual faculty to indicate sources of assistance in making such improvements will hopefully be explored. Some consideration has been given to shaping the current device to a greater degree, toward functioning as a general barometer with respect to the major elements of systematic instruction while developing a series of shorter, more focused diagnostic instruments with which to pinpoint the specific facets of the instructional system need remediation. For example, one such diagnostic device might dwell entirely on the suitability and availability of the learning materials used in the course and contain items related to readability, cultural bias, accessibility and the like. The overall goal of such explorations would however remain what it has always been; to develop a program of collecting and using student perceptions of instructional experiences which is based on the model of systematic instruction adopted by College of the Mainland.

APPENDIX A

REPRESENTATIVE COMPUTER PRINTOUT OF RESULTS

EXPLANATIONS OF SYMBOLS

X = Arithmetic Mean
SD = Standard Deviation
WT = Item Weight
NA = Not Applicable
SG = Significantly High
SL = Significantly Low
LO
AV = Average
HI = Confidence Limits (We may assume with 95% surity that the "True" mean lies within these limits)

NUMBER IN EACH CLASSIFICATION

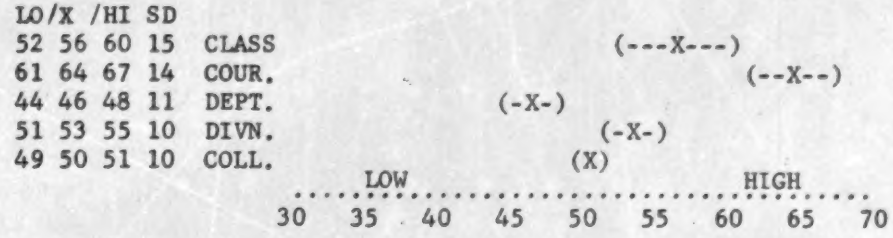
DIVN.	DEPT.	COUR.	CLASS	TOTL.
4	30	130	400	12000
	5	60	100	3000
		10	50	1550
			3	90
				30

S.A.S.I. RESULTS

INST.-College of the Mainla
DIVN.-Industrial Education
DEPT.-Welding
COUR.-Arc Welding 101
SECT.-.01

FACTOR - LEARNING MATERIALS

DISPLAY OF MEANS AND CONFIDENCE LIMITS



SUMMARY OF COMPARISONS

COMPARED TO	THIS CLASS RATES	THE COURSE RATES	THE DEPARTMENT RATES	THE DIVISION RATES
THE COURSE	-BELOW AVERAGE-			
DEPARTMENT	-ABOVE AVERAGE-	- HIGH -		
DIVISION	- AVERAGE -	- HIGH -	-BELOW AVERAGE-	
COLLEGE	-ABOVE AVERAGE-	- HIGH -	-BELOW AVERAGE-	-ABOVE AVERAG

	RATE	MEAN	SD.	WT.	%-NA	ITEM		
ITEM	AV	3.3	1.0	.82	3 HI	21	PRINTED MATERIALS ARE READABLE.	ITEM
BY	AV	3.3	.9	.80	3 HI	3	I LIKE THE LEARNING MATERIALS THAT ARE BEING USED.	BY
ITEM	AV	3.0	1.0	.70	5 HI	10	THE AUDIO-VISUAL MATERIALS FOR THIS COURSE ARE HELPFUL.	ITEM
ANALYSIS	AV	2.6	.9	.30	5 AV	36	STUDENTS HAVE A CHOICE OF LEARNING MATERIALS.	ANALYSIS
	AV	2.4	.8	.25	6 AV	55	NECESSARY MATERIALS ARE USUALLY AVAILABLE.	
	AV	1.8	1.0	.75	7 LO	42	THE TEXTBOOK IS SATISFACTORY.	
	AV	1.3	.9	.66	4 LO	28	MATERIALS SEEM TO BE PRESENTED ON MY LEVEL.	

REFERENCES

American Association of University Professors Committee C on Teaching, Research and Publication. Statement on teaching evaluation. AAUP Bulletin, 1974, 60(2), 166-170.

Bannister, J., & Others. Evaluating college teaching. Curriculum Reporter Supplement, December 1961, (1). San Jose State College, California. Eric Document No. ED022450.

Bendig, A.W. An inverted factor analysis student-rated introductory psychology instruction. Journal of Experimental Education, 1953, 21, 333-336.

Bryson, R. Teacher Evaluations and Student Learning. Journal of Educational Research, 1974, 68: 12-14.

Centra, J.A. Problems of evaluating teaching. Current Issues in Higher Education, Association of Higher Education, Washington, D.C., 1966.

Centra, J.A. The Student As Godfather? The Impact of Student Ratings on Academia. Educational Researcher, 1973, 3(9), 4-8.

Centra, J.A. The Student Instructional Report: Its Development and Use. SIR Report No. 1, Princeton, New Jersey, Educational Testing Service, 1972.

Centra, J.A. The Utility of Student Ratings for Instructional Improvement. Project Report 72-16. Princeton, New Jersey, Educational Testing Service, 1972.

Coffman, W. E. Determining students' concepts of effective teaching from their ratings of instructors. Journal of Educational Psychology, 1954, 45, 277-286.

Cohen, S. A., & Berger, W. G. Dimensions of students' ratings of college instructors underlying subsequent achievement on course examinations. Proceedings of the 178th Annual Convention of the American Psychological Association, 1970, 5, 605-606.

Costin, F. A graduate course in the teaching of psychology: Description and Evaluation. Journal of Teacher Education, 1968, 19, 425-432.

Costin, F.; Greenough, W. T.; & Menges, R. J. Student Ratings of College Teaching: Reliability, Validity, and Usefulness. Review of Educational Research, 1975, 46, 461-470.

Crittenden, K. S., Norr, J. L., & LeBailly, R. K. Size of university classes and student evaluations of teaching. Journal of Higher Education, 1975, 46, 461-470.

Crittenden, K. S., Norr, J. L. Some remarks on student ratings: The validity problem. American Educational Research Journal, Fall, 1975, 12(4), 429-434.

Crittenden, K. S., & Norr, J. L. Student values and teacher evaluation: A problem in person perception. Sociometry, 1973, 36, 143-151.

Crawford, P. L., & Bradshaw, H. L. Perception of characteristics of effective university teachers: A scaling analysis. Educational and Psychological Measurement, 1968, 28, 1079-1085.

Cronbach, L. J., & Furby, L. How we should measure "change" - or should we? Psychological Bulletin, 1970, 74, 68-90.

Deegan, W. L. Should students evaluate faculty? Community and Junior College Journal, Oct. 1972, 43(2), 25-26.

Dobbins, J. O. Jr. Review of Course Evaluation by Students at College of the Mainland. Austin, Texas: College Associates, 1975.

Eagle, Norman. Validity of Student Ratings: A Reaction. Community and Junior College Journal, 1975, 6-8.

Ebel, K. E. The recognition and evaluation of teaching. Salt Lake City, Utah: Project to Improve College Teaching, 1970.

French-Lazovik, G. Predictability of students' evaluations of college teachers from component ratings. 1974, 66(3), 373-385.

Frey, P. W. Student evaluation: Technical comment. Science, 1975, 187, 557-558.

Frey, P. W.; Leonard, D. W.; & Beatty, W. W. Student Ratings of Instruction: Validation Research. American Educational Research Journal, 1975, 12(4), 435-443.

Frey, P. W. Student ratings of teaching: Validity of several rating factors. Science, 182: 83-85, 1973.

Frey, P. W. The Ongoing Debate: Student Evaluation of Teaching. Change, Feb., 1974, p. 47ff.

Gage, N. L. The appraisal of college teaching. Journal of Higher Education, 1961, 32, 17-22.

Gessner, P. K. Evaluation of instruction. Science, 1973, 180, 566-569.

Gessner, P. K. Student evaluation: Technical comment. Science, 1975, 187, 558-559.

Gustad, J. W. Evaluation of teaching performance: Issues and possibilities. In C. Lee (Ed), Improving College Teaching. Washington, D.C.: American Council on Education, 1967.

Gustad, J. W. Policies and practices in faculty evaluation. Educational Research, 1961, 42, 194-211.

Guthrie, E. R. The evaluation of teaching: A progress report. Seattle: University of Washington, 1954.

Guthrie, E. R. The evaluation of teaching. Educational Record, 1949, 30, 109-115.

Hartley, E. L. & Hogan, T. P. Some additional factors in student evaluation of courses. American Educational Research Journal, 1972, 9, 241-250.

Harvey, J. N.,& Barker, D. G. Student evaluation of teaching effectiveness. Improving College and University Teaching, 1970, 18, 275-278.

Heilman, J. D., & Arementrout, W. D. The rating of college teachers on ten traits by their students. Journal of Educational Psychology, 1936, 27, 199-216.

Heinz, E. Student opinion survey, Grossmont College, 1967. Grossmont College, El Cajon, California. Eric Document No. ED017233.

Hildebrand, M., & Wilson, R. C. Effective university teaching and its evaluation. Berkeley, Calif.: Center for Research and Development in Higher Education, University of California, 1970

Hildebrand, M., Wilson, R. C., & Dienst, E. R. Evaluating University Teaching, Berkeley: Center for Research and Development in Higher Education, 1971.

Isaacson, R. L.; McKeachie, W. J.; Milholland, J. E.; Lin, Y. G.; Hofeller, M.; Baerwaldt, J. W.; & Zinn, K. L. Dimensions of student evaluations of teaching. Journal of Educational Psychology, 1964, 55, 344-351.

Johnson, D. J., & Leider, F. E. Comparison of students attitude on performance based on traditional teaching methods. Eric Document No. 096309.

Lathrop, R. G. Unit factorial ratings by college students of courses and instructors. Ohio State College, California, 1968. (Mimeographed)

Lovell, G. D., & Haner, C. F. Forced choice applied to college faculty rating. Educational and Psychological Measurement, 1955, 15, 291-304.

Maslow, A. H., & Zimmerman, W. College teaching ability, scholarly activity, and personality. Journal of Educational Psychology, 1956, 47, 185-189.

McKeachie, W. J., & Solomon, D. Student ratings of instructors: A validity study. Journal of Educational Research, 1958, 51, 379-382.

McKeachie, W. J., Lin, Y., & Mann, W. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 1971, 8, 435-445.

McKeachie, W. J., & Lin, Y. G. Use of student ratings in evaluation of college teaching. Final Report. Ann Arbor, Michigan, March, 1975.

Menges, R. J. The new reporters: Students rate instruction. In C. R. Pace (Ed), New Directions in Higher Education: Evaluating learning and teaching. San Francisco: Jossey-Bass, 1973.

Morsh, J. E., Burgess, G. G., & Smith, P. N. Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 1956, 47, 79-88.

Nelson, D. E., & Dobbins, J. O. Jr. Student Assessment of Systematic Instruction: A Users Manual. Texas City, Texas, College of the Mainland, December, 1975.

Owen, P. H. Some Dimensions of College Teaching: An explanatory Study Using Critical Incidents and Factor Analyses of Student Ratings. Unpublished Doctoral Dissertation, University of Houston, 1967.

Rashdall, H. The Universities of Europe in the Middle Ages, Volume I. London: Oxford University Press, 1936. (From a new edition, 1964 Oxford University Press, F. M. Powicke and A. B. Emden (Eds.).)

Rayder, N. F. College student ratings of instructors. Eric Document No. Ed021527.

Rayder, N. F. College student ratings of instructors. Journal of Experimental Education, 1968, 37, 76-81.

Remmers, H. H. Manual, The Purdue Rating Scale for Instruction. (Revised Edition) West Lafayette, Indiana: University Book Store, 1960.

Rezler, A. G. The influence of needs upon the student's perception of his instructor. The Journal of Educational Research, 1965, 58, 282-286.

Rodin, N., Frey, P. W., & Gessner, P. K. Student evaluation. Science, 1975, 187, 555-559.

Rodin, M., & Rodin, B. Student evaluation of teachers. Science, 177: 1164-1166, 1972.

Ronan, W. W. Evaluating college classroom teaching effectiveness. Promoting Research in Educational Practice, Report No. 34. Atlanta, Georgia, 1973.

Rosenshine, B., Cohen, A., & Furst, N. Correlates of student preference ratings. Journal of College Student Personnel, 1973, 14, 269-272.

Silberman, M. L., & Allender, J. S. A Semiprojective Technique for Assessing Students' Reactions to College Classes. Journal of Higher Education, 1974, 65(6), 450-457.

UNIVERSITY OF CALIF.
LOS ANGELES

JUL 9 1976

CLEARINGHOUSE FOR
JUNIOR COLLEGES

Slobin, D. Y., & Nichols, D. G. Student rating of teaching. Improving College and University Teaching, 1969, 17, 244-248.

Smith, A. A. College teachers evaluated by students. Sociology and Social Research, 1944, 28, 471-478.

Spencer, R. E., & Aleamoni, L. M. A student course evaluation questionnaire. Journal of Educational Measurement, 1970, 7, 209-210.

Spencer, R. E. The Course evaluation questionnaire: Manual of interpretation. Office of Institutional Research, University of Illnois, 1965.

Sockloff, A. L., & Deubler, V. T. The construction of the faculty and course evaluation instrument. Research Report. Philadelphia, Pa.: Temple University Testing Bureau, March, 1971.

Sorotnik, K. A. Introduction to matrix sampling for the practitioner. In W. F. Popham (Ed.), Evaluation in education: Current applications. Berkeley, Calif.: McCutchen, 1974, 451-529.

South, J. D., & Others. The relationship of teacher role orientation and perceived teaching effectiveness. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., March 31-April 3, 1975. Eric Document No. ED10446.

Sullivan, A. M., & Skanes, G. R. Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 1974, 66, 584-590.

Twa, R. J. Student and instructor expectations of community college instructors. Unpublished doctoral dissertation, 1970. Eric Document No. ED046383.

Vandervert, L. R. Student evaluation of instruction: Some theoretical considerations and a proposal. Paper presented at Meeting of the Washington State Community College, District 17 Board of Trustees, March, 1974. Eric Document No. ED093394.

Walker, B. D. An investigation of selected variables relative to the manner in which a population of unior college students evaluate their teachers. Dissertation Abstracts, 1969, 29(9-B), 3474.

Webb, W. B., & Nolan, C. Y. Student supervision and self ratings of instructional proficiency. Journal of Educational Psychology, 1955, 46, 42-46.

Werdell, P. R. Course and Teacher Evaluation (2nd Edition). Washington, D.C. United States National Students Association, 1967.

Young, J. I., & Van Mondfrans, A. P. Psychological implications of competency based education. Educational Technology, 1972, 12(11), 15-18.